91e Année - N° 796 Publication Mensuelle Juillet/Août/Septembre 1997

Bulletin de l'Union des Physiciens

Association des professeurs de Physique et de Chimie

Régression linéaire et incertitudes expérimentales

par Daniel BEAUFILS
Institut National de Recherche Pédagogique
Département Technologies Nouvelles
91, rue Gabriel Péri - 92120 Montrouge
et Hélène RICHOUX
Lycée Marcel Pagnol
Avenue de la Terrasse - 91200 Athis-Mons

RÉSUMÉ

L'utilisation de la régression linéaire dans l'analyse de résultats expérimentaux suscite toujours de nombreuses discussions conduisant parfois à l'abandon même de la méthode. Mais «l'erreur» vient parfois uniquement de ce que l'on n'utilise pas tous les résultats du calcul : issue d'un critère quadratique qui contient des hypothèses sur les incertitudes, la régression ne conduit pas en effet seulement aux valeurs de deux paramètres, mais aussi à leur propre incertitude (donc aux chiffres significatifs) et à l'incertitude même des mesures portées en ordonnée. Une telle exploitation permet alors, sur le plan scientifique, de résoudre certaines situations sinon indécidables (nullité de l'ordonnée à l'origine, point aberrant, etc.). Sur le plan didactique, l'explicitation permet de clarifier l'articulation des activités de modélisation et de mesurage et d'introduire ainsi légitimement la prise en compte des incertitudes expérimentales. Une progression coordonnée avec l'enseignement de mathématiques en terminale scientifique est donnée à titre d'exemple.

INTRODUCTION

L'utilisation de la régression linéaire dans l'analyse de résultats expérimentaux suscite toujours de nombreuses discussions conduisant parfois à l'abandon même de la méthode pour des raisons «mathématiques» ou «physiques» : dans tel cas les conditions sur les incertitudes des grandeurs portées sur les axes (incertitude constante sur la grandeur portée en ordonnée et négligeable sur celle portée en abscisse) ne sont pas respectées, dans tel autre cas, la linéarité de telle relation étant connue, la détermination de la pente ne doit pas se faire par régression mais en choisissant un point donnant la plus faible incertitude sur le calcul du rapport (en général le point le plus «éloigné» de l'origine).

On peut d'ailleurs noter à ce propos qu'une partie des difficultés en classe de lycée vient de ce que, pour des raisons d'économie de temps, on est amené à vouloir atteindre plusieurs objectifs à la fois : la découverte d'une corrélation entre deux grandeurs, la validation d'un modèle linéaire (ou affine) et la détermination du coefficient directeur qui est le paramètre «central» du cours ou du T.P. Notre propos se situe cependant d'abord sur le plan scientifique : la course aux résultats évoquée ici, serait moins discutable si elle s'appuyait sur une réelle exploitation des mesures et de la méthode.

La régression linéaire ne fournit en effet pas seulement deux valeurs, mais tout l'ensemble des résultats issus de ce calcul statistique : incertitude sur chacun des paramètres, incertitude sur les valeurs portées en ordonnée et coefficient de corrélation, notamment.

Ainsi, par exemple, le tableur EXCEL (© Microsoft) permet-il de calculer une régression linéaire avec la fonction DROITEREG ou TENDANCE. Le résultat du calcul apparaît dans un tableau 1 où figurent l'erreur-type sur y (s_y), sur a (s_a) et sur b (s_b), ainsi que le coefficient dit de détermination (r^2).

a	b
s_a	s_b
r^2	s_y

Tableau 1

Or, bien souvent, l'affichage des valeurs des deux paramètres *a* et *b* est le seul résultat retenu ce qui, suivant les cas, conduit à une largesse dans l'accord modèle/mesures ou, plus grave, à des «conclusions» portant leur propre contradiction comme, par exemple, le cas de la détermination d'une accélération de la pesanteur égale à 9.93 m/s^2 (si les trois chiffres sont maintenus, ou bien il y a un effet de pesanteur local (!) ou bien il y a un biais dans le protocole expérimental ; sinon...).

«L'erreur» dans l'utilisation de la régression linéaire vient donc parfois uniquement du fait que *l'on n'utilise pas tous les résultats du calcul*. Notre objectif ici est de donner quelques indications complémentaires sur ce sujet et de montrer pourquoi la prise en compte de ces éléments[1] est non seulement nécessaire à une utilisation valide de la régression mais également particulièrement féconde sur le plan de l'analyse des incertitudes expérimentales. Dans une première partie nous rappelons les résultats «habituels» concernant la régression, puis dans une seconde partie nous attirons l'attention sur la démarche, qui respectée jusqu'au bout, permet de prendre en compte sérieusement les incertitudes. Enfin nous terminerons en proposant quelques éléments de réflexion relatifs à l'introduction d'une prise en compte des incertitudes expérimentales et de leur traitement statistique en classe de physique de lycée.

LES «CLASSIQUES» DE LA RÉGRESSION LINÉAIRE

Les «formules» de la régression

Le titre peut sans doute surprendre, de même que le mode d'introduction. Mais il s'agit d'une part de faire référence au «point d'entrée» des élèves et étudiants - la calculette - et d'autre part d'indiquer très clairement que ces formules que l'on trouve couramment ne sont pas au centre de notre propos, mais *au point de départ*...

Pour une série de N couples (x_i, y_i), les coefficients de la droite de régression $y = ax + b$ sont :

$$\boldsymbol{a} = \frac{N\ \Sigma x_i y_i - \Sigma x_i\ \Sigma y_i}{N\ \Sigma x_i^2 - \left(\Sigma x_i\right)^2} \qquad \boldsymbol{b} = \frac{\Sigma y_i\ \Sigma x_i^2 - \Sigma x_i\ \Sigma x_i y_i}{N\ \Sigma x_i^2 - \left(\Sigma x_i\right)^2}$$

Les conditions d'application

Les conditions d'utilisation de la régression linéaire comme méthode d'optimisation stipulent *que l'incertitude doit être négligeable pour la grandeur portée en abscisse et constante pour celle portée en ordonnée.*

Ces conditions sont essentielles sous peine de conduire à des déterminations plus «mauvaises» que celles obtenues avec une règle et un crayon. C'est évidemment le cas lorsque l'incertitude de la grandeur portée en abscisse est aussi grande (sinon plus) que celle de la grandeur portée en ordonnée ; mais c'est aussi le cas de tous les changements

1. Nous ne retiendrons pas le coefficient de corrélation qui n'est généralement pas pertinent en physique : l'existence de la corrélation n'est pas en doute (contrairement à d'autres domaines d'utilisation de ces calculs).

de variables[2] : en effet si, par chance, l'incertitude sur la grandeur mesurée G est constante, alors *on est certain que la seconde condition ne peut être vérifiée* si l'on porte en ordonnée G^2, 1/G, log(G/Go), etc.

Ces deux aspects sont fondamentaux, non seulement sur le plan de la rigueur scientifique, mais sur le plan de la clarté didactique. Ne pas expliciter le fait que la grandeur portée en abscisse a une incertitude négligeable, est laisser passer l'idée qu'il en est généralement ainsi, d'autant plus que la procédure classique renforce le trait : on porte souvent une valeur de Y mesurée pour une valeur de X *«donnée»*. Par ailleurs, l'idée de faire passer une droite au mieux parmi un ensemble de points, c'est-à-dire implicitement de façon que tous les points soient proches de la droite, repose bien sur l'hypothèse qu'ils sont tous de même importance ; tout élève comprend bien sinon qu'*il n'y aurait aucun sens à vouloir faire passer la droite aussi près d'un «point incertain» que d'un autre déterminé avec une grande précision*[3].

L'origine de ces formules : l'écart quadratique

Notre approche, inverse de celle pratiquée habituellement, conduit alors naturellement à expliquer d'où viennent ces formules et ces conditions particulières. L'objectif est de suivre une démarche où l'on fait surgir les questions avant d'y répondre et l'on peut alors montrer *l'évidence* qui se cache derrière certains énoncés.

L'écart quadratique et les moindres carrés

L'idée de base est de trouver une mesure de l'écart entre un modèle mathématique et des données expérimentales, et d'utiliser cette mesure comme critère d'adéquation : le meilleur modèle sera alors celui qui minimise cette mesure[4].

Le choix le plus simple au niveau des calculs consiste à mesurer l'écart entre la courbe théorique et les points, en considérant pour chaque point la différence de coordonnées $(y_{th} - y_{exp})$ entre le point expérimental et le point théorique de même abscisse calculé avec l'équation du modèle ($y = f(x)$) (figure 1). L'écart quadratique moyen *(J)* que l'on considère alors est la somme (moyennée sur N mesures) des écarts calculés pour chaque point et élevés au carré :

$$J = \sum (y_{i_{th}} - y_{i_{exp}})^2 / N \qquad \text{avec : } y_{i_{th}} = f(x_{i_{\mathrm{exp}}})$$

2. Parfois appelés «anamorphoses»...
3. Pour le détail des analyses et les compléments d'explication, voir le cours d'Yves CORTIAL (UdP-INRP, 1995).
4. Voir «Recherche de modèles expérimentaux» de J.-C. TRIGEASSOU (1988) et «Identification de modèles paramétriques à partir de données expérimentales», E. WALTER et L. PRONZATO, Masson (1994).

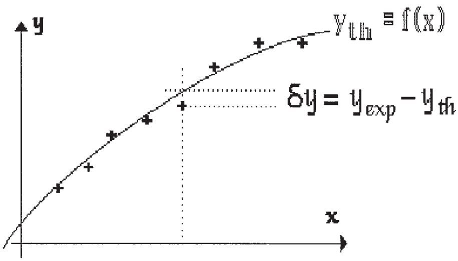

Figure 1 : L'idée d'écart (compé suivant y).

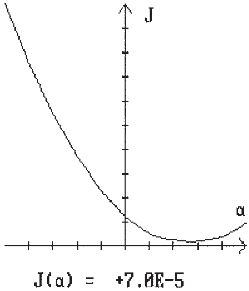

Figure 2 : Exemple de variation de l'écart en fonction d'un paramètre du modèle.

Cet écart, ainsi défini pour toute fonction mathématique, dépend évidemment de la valeur des paramètres entrant dans l'expression de cette dernière. Si *a* désigne un de ces paramètres, l'écart peut être considéré comme une fonction *J(a)* dont la valeur est d'autant plus grande que celle de *a* s'écarte de la «bonne» valeur ; en d'autres termes, *J(a)* passe par un minimum pour la «meilleure» valeur (voir exemple tiré du logiciel Chute[5], figure 2). La recherche de la valeur d'un paramètre par une telle méthode est appelée *méthode des moindres carrés* (simple).

Les conditions d'utilisation inscrites dans l'expression du critère

Si ce critère est particulièrement simple, il contient *à l'évidence* des conditions d'utilisation. En effet, l'expression choisie ne fait intervenir que l'écart compté sur l'axe «des y» ; ceci signifie que *l'incertitude* considérée n'est *que celle qui concerne y* (ou, en d'autres termes, que *l'incertitude sur la grandeur portée en abscisse est négligeable).* Par ailleurs, la sommation considère *tous les points de la même manière* ; cela signifie que l'incertitude est considérée comme étant la même pour tous ou, en d'autres termes, que *l'incertitude est indépendante de la grandeur portée en abscisse.*

La minimisation

On comprend bien que par itérations successives, en faisant varier le paramètre à déterminer, on puisse aboutir au minimum[6]. Lorsque le modèle est linéaire par rapport

5. Logiciel Chute (INRP-Jeulin) : étude du mouvement de chute verticale ; voir aussi Plan (INRP-Jeulin) et Table (INRP-CNDP), logiciels dans lesquels ces outils de calcul et de représentation graphique ont été implantés ; voir aussi Actilab (Jeulin) qui affiche l'évolution de l'écart quadratique au cours du calcul d'optimisation.
6. Différentes méthodes peuvent être utilisées pour assurer une convergence la plus rapide possible ; voir la documentation du logiciel Regressi (auteur J.-M. Millet, édition Micrelec) ; voir aussi l'annexe des actes de l'université d'été INRP-UdP, 1993.

au paramètre (mais quelconque par rapport à x), la courbe $J(a)$ représentée sur la figure 2 est évidemment une parabole ($f(x)$ étant linéaire en a, la somme des carrés fait apparaître un terme en a^2) et son minimum peut être calculé analytiquement.

La difficulté vient en général de ce que le modèle comporte plus d'un paramètre, et c'est le cas de la régression dite «linéaire»[7]. Néanmoins les coefficients a et b du modèle $y = ax + b$ peuvent être calculés analytiquement.

Le minimum de J (noté J_{min}), fonction de a et b, se détermine par l'annulation des dérivées partielles :

$$\partial J / \partial a = 0 \qquad \text{et} \qquad \partial J / \partial b = 0$$

Le calcul des dérivées puis la résolution du système de deux équations à deux inconnues ne présente aucune difficulté particulière[8] et conduit aux expressions données précédemment.

L'INFORMATION SUR LES INCERTITUDES

L'estimation du sigma-y

La figure 3 montre qualitativement comment se place la droite lorsqu'on a minimisé l'écart quadratique entre celle-ci et l'ensemble des points. L'écart résiduel (qui reste lorsqu'on a fait «au mieux») doit alors être lu comme l'écart irréductible *dû aux incertitudes expérimentales* elles-mêmes.

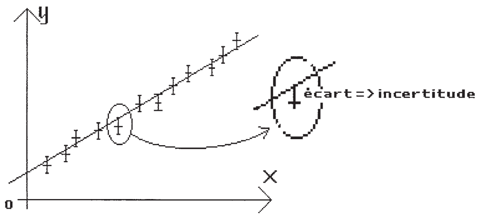

Figure 3 : Représentation des écarts résiduels (dans le cas d'un modèle linaire).

En d'autres termes, on comprend bien que *si* l'on considère tous les points de la même manière et *si* l'on impose le type de modèle, *alors* l'écart résiduel minimum est la représentation de l'incertitude globale de l'acquisition. Cette dernière formulation lève l'ambiguïté de la démarche sous-jacente et évite toute discussion stérile sur le fait que

7. Expression malheureuse pour les élèves de lycée...
8. Sauf pour des élèves de lycée qui ne connaissent pas les fonctions à deux variables.

l'on pourrait trouver un modèle qui donne un écart plus faible, y compris un polynôme de degré N – 1 qui passerait par tous les points...

À ce propos il faut bien voir que chaque mesure peut être obtenue avec une bonne précision (l'appareil de mesure est précis et exact) mais que l'incertitude peut provenir de la répétabilité de l'acte de mesure : la dispersion que l'on observe de part et d'autre de la droite est bien celle qui provient des fluctuations d'une mesure à l'autre et peut avoir des causes multiples et aléatoires (fluctuation d'une alimentation, parasites extérieurs, mauvaise reproductibilité de conditions initiales, etc.).

On comprend bien aussi que l'information sur les incertitudes est contenue dans la valeur minimale de l'écart quadratique. Il n'est pas dans notre propos de prétendre que la démonstration est évidente, mais de faire remarquer que l'idée d'une relation entre l'estimation d'un écart-type et le minimum de l'écart quadratique est d'autant moins surprenante que ces deux grandeurs relèvent du même procédé : la sommation des carrés des écarts à une valeur moyenne (la moyenne au sens habituel pour l'écart-type, la valeur donnée par la droite moyenne dans l'écart quadratique) :

$\sigma_y{}^2 = \sum (y_i - m)^2 / N$ (avec $m = \sum y_i / N$)	$J_y = \sum (y_{i_{th}} - y_{i_{exp}})^2 / N$
Expression de l'écart-type	Expression de l'équart quadratique

Dans le cas de la régression linéaire, la relation entre la valeur minimale de J_{min} et l'estimation s_y de σ_y est particulièrement simple[9] :

$$s_y^2 = \frac{N}{N-2} J_{min}$$

De sorte que, à l'issue d'une régression linéaire, non seulement la droite peut être tracée, mais également les barres d'incertitudes correspondant, par exemple, à un écart de $\pm 3\, s_y$; celles-ci, toutes identiques, indiquent alors bien à la fois *les hypothèses de départ* et *les conséquences que le choix du modèle linéaire impose en retour*.

L'incertitude sur les paramètres a et b

On conçoit dès lors que les incertitudes sur les y_i entraînent une incertitude sur les valeurs des paramètres. En quelque sorte, et conformément à ce que l'on expliquait au précédent paragraphe, *ce que l'on trouve n'est pas exactement ce que l'on cherche*. Les valeurs rendues pour *a* et *b* sont calculées de façon «exacte» et doivent être données avec le même nombre de chiffres significatifs que celui utilisé pour chaque valeur mesurée (qui

9. Voir J.-C. TRIGEASSOU (1988) : chapitre 11, § 4.2.

peut être important si l'instrument de mesure est précis). Mais ce qui nous intéresse, ce ne sont pas *ces valeurs* de la pente et de l'ordonnée à l'origine, mais les valeurs des paramètres *a* et *b* du modèle, dont ce calcul issu de ces mesures ne donne qu'une *estimation*.

Là encore, si l'on imagine bien que par propagation des incertitudes[10] il est possible de calculer analytiquement les estimations s_a et s_b des écarts-types σ_a et σ_b affectant les paramètres *a* et *b*, il s'agit surtout de comprendre que la régression linéaire donne bien «en retour» une information sur l'incertitude des estimations des paramètres. À l'issue d'un tel calcul on *doit* donc écrire :

$$a = valeur\ 1 \pm \ incertitude\ 1$$

$$b = valeur\ 2 \pm \ incertitude\ 2$$

Notons enfin, que les valeurs des incertitudes alors calculées, sont elles-mêmes des estimations, c'est-à-dire des valeurs accompagnées d'une incertitude. En conséquence, il conviendra d'arrondir les valeurs des incertitudes, puis de tenir compte de cet arrondi pour écrire la valeur avec le bon nombre de chiffres significatifs. Ainsi, un calcul qui donnerait, par exemple :

$$a = 50.975 \pm 0.166 \text{ (unité)}$$

doit se transcrire : $a = 51.0 \pm 0.2$ (unité)

Des informations particulièrement utiles

Ainsi que nous venons de le voir, utiliser la régression linéaire *impose* que l'on fasse l'hypothèse requise sur les incertitudes et que l'on aille jusqu'au bout du raisonnement, c'est-à-dire accepter l'estimation des incertitudes de la grandeur portée en ordonnée qui en résulte, ainsi que celle sur l'estimation des paramètres. Une méthode de régression ne devrait jamais s'utiliser en omettant ces indicateurs précieux. On peut même imaginer (et réaliser) un algorithme qui, à l'issue d'un calcul, trace la droite et la complète par la représentation des barres correspondantes tracées à $\pm 3\, s_y$ autour des points expérimentaux, et affiche l'équation avec les valeurs des paramètres données avec le bon nombre de chiffres significatifs[11] !

10. Voir de nombreux articles sur les incertitudes ; voir aussi, D. Beaufils et R. Journeaux (1990).
11. Savoir s'il vaut mieux, sur le plan didactique, avoir un outil qui donne des affichages tenant compte du sigma y (mais qui cache le problème) ou garder des affichages «à discuter», reste une question à étudier.

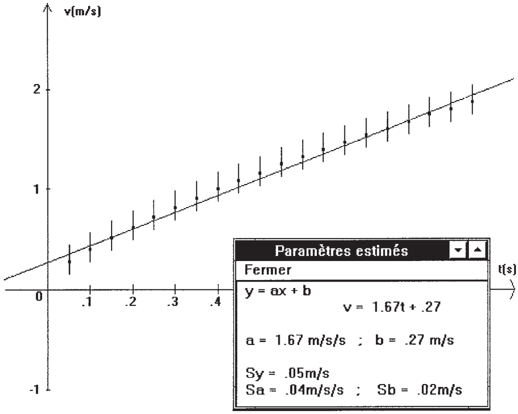

Figure 4 : Exemple de représentation graphique avec affichage «complet» issue d'une régression linéaire[12].

Au-delà du fait qu'il est évidemment nécessaire de donner un résultat déduit de l'expérience sous forme d'un nombre *et de son incertitude*, ces informations sont essentielles pour résoudre quelques cas sinon indécidables.

Pour avoir un discours scientifiquement correct

Tous les exemples d'utilisation de la régression linéaire avec comme simple résultat de noter les deux valeurs affichées par le logiciel ou la calculette sont donc critiquables. En l'absence d'information sur ces valeurs, on ne peut en effet rien conclure, et toute déclaration sur la bonne réalisation de l'expérience qui a donné un résultat très proche de celui attendu est purement rhétorique ! Si, par exemple, le seul résultat pour la valeur de l'accélération de la pesanteur terrestre est $9.8\ m.s^{-2}$, alors il convient de *ne rien conclure* ; il faut en effet savoir si, par ailleurs, l'incertitude n'est pas de $\pm 1\ m.s^{-2}$... Dans ce cas, il vaut mieux considérer une autre détermination ayant donné $9.7\ m.s^{-2}$ avec une incertitude de $\pm 0.1\ m.s^{-2}$...

Pour résoudre la question du passage par l'origine...

Dans de nombreuses circonstances[13] la relation entre deux grandeurs est linéaire (au sens strict). Cependant, l'utilisation de la régression sur l'ensemble de points conduit inévitablement à une valeur de l'ordonnée non strictement nulle. Comment alors

12. Logiciel Traceur ; écrit en Visual-Basic par l'auteur ; diffusion UdP (disquette n° 11).
13. En physique !

passer au modèle linéaire sans information sur les incertitudes ? *A contrario*, l'affichage de la valeur avec son incertitude permet de trancher. Ainsi un affichage «brut» tel que :

$$b = 1,81 \pm 9.44 \text{ (unité)}$$

permet-il d'accepter un modèle linéaire (b = 0).

La question du point aberrant, de la validité du modèle, etc.

Si les barres d'incertitudes qui apparaissent sont notablement supérieures à celles qui correspondraient à l'incertitude estimée à partir des appareils ou du protocole expérimental, plusieurs hypothèses doivent être envisagées :

– les estimations/hypothèses sur les incertitudes expérimentales (sur x et/ou y) sont erronées,

– un (ou plusieurs) point(s) doivent être considérés comme aberrants,

– le modèle linéaire ne convient pas.

L'exemple de la figure 5 est obtenu en prenant à partir d'une série de huit points dont le cinquième n'est pas dans l'alignement des autres. Le fait de maintenir tous les points, le modèle et la méthode d'analyse, impose alors les valeurs des incertitudes. La représentation met en évidence le «problème» et doit entraîner l'abandon d'une (au moins) des trois hypothèses. Chercher l'erreur renvoie alors à de nouvelles considérations sur le modèle, sur la méthode d'analyse, sur les mesures. Autant de questions essentielles dans une pratique scientifique.

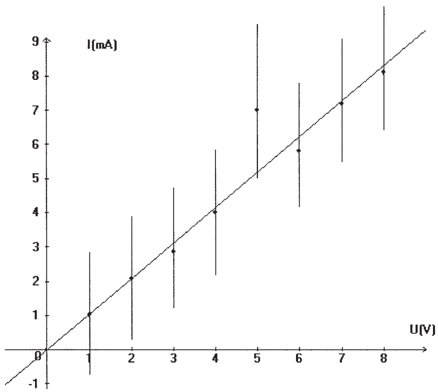

Figure 5 : Le maintien de tous les points (de façon équivalente) et du modèle linéaire impose que l'on accepte de telles incertitudes...

QUELQUES IDÉES-FORCES POUR CONCLURE

En guise de conclusion, on peut dire qu'il y a, au-delà des expressions mathématiques, quelques préceptes essentiels à une bonne utilisation de la régression linéaire :

– ne pas réduire la régression linéaire à la seule détermination des deux paramètres de l'équation de la droite (mais en tirer les informations sur les incertitudes),

– ne pas confondre «test de modèle» et «estimation de paramètres» : dans nombre de cas, la régression linéaire est utile pour estimer l'adéquation du modèle ; ce qui est important alors c'est la valeur de l'incertitude s_y que cela impose en retour ; par contre, si l'objectif est la détermination de tel ou tel paramètre, il est en général intéressant de prendre une autre méthode, tenant compte de la validité *acceptée* du modèle : il est en effet illogique de calculer la pente en considérant des points près de l'origine qui viendraient *détériorer* une bonne détermination à partir de points «éloignés»,

– éviter «l'anamorphose» : hormis le cas de validation d'une hypothèse quant à la forme d'un modèle (en x, x^2, log(x), etc.) pour laquelle le changement de variable permet de se ramener à une allure facile à «tester» («test du modèle»), *l'anamorphose est par nature incompatible avec les conditions d'application de la régression linéaire*.

Deux remarques enfin :

– tous ces éléments développés à propos de la régression linéaire valent en fait pour toute modélisation par la méthode des moindres carrés simple (non pondérée). Ainsi que nous l'avons indiqué, la fonction *f(x)* qui intervient dans l'écart quadratique est quelconque,

– si l'on se penche un peu sur la question, en particulier à propos des points aberrants, on finit par comprendre que l'écart entre un point et la courbe théorique n'a, dans le cas général, pas de valeur en soi : c'est sa valeur *rapportée à l'incertitude* du point qui est intéressante, savoir si c'est dans la «norme» ou non... Rapporter l'écart à l'incertitude, c'est-à-dire estimer $\left(y - f(x)\right)^2 / \sigma_y^2$ (donc un χ^2), voilà qui est bien naturel *in fine*...

Tous ces éléments ne sont pas que des idées «théoriques». Non seulement elles nous apparaissent comme essentielles au niveau de la vigilance scientifique, mais elles font l'objet d'une expérimentation dans le cadre d'une recherche INRP centrée sur l'utilisation des instruments informatisés d'investigation scientifique en physique : le schéma de la progression expérimentée sur les cours de physique et de mathématiques est donné en annexe 2.

BIBLIOGRAPHIE

Note : La bibliographie relative aux incertitudes est très importante, y compris sur le seul plan scientifique ; nous ne citons ici que quelques articles et ouvrages d'accès facile et qui contiennent de nombreuses références.

– D. BEAUFILS et R. JOURNEAUX : *«Physique et informatique, une approche programmatique»*, Versailles : CARFI, 1990, 83 pages.

– Y. CORTIAL : *«La prise en compte des incertitudes dans les méthodes d'optimisation»*, in actes de l'Université d'été «outils informatiques d'investigation scientifique dans l'enseignement des sciences physiques», UdP, pp. 61-96, 1995.

– Y. CORTIAL : *«À propos de la méthode des moindres carrés»*, BUP n°725, juin 1990, pp. 769-791.

– R. JOURNEAUX : *«La régression linéaire et ses conditions d'application»*, BUP n° 752, mars 1993, pp. 353-369.

– G. TROUILHET, R. CULOS, J.-C. TRIGEASSOU, D. BEAUFILS et J.WINTHER : *«Acquisition et analyse de données»*, Paris : UdP-INRP, 1990, 128 pages.

– J.-C. TRIGEASSOU : *«Recherche de modèles expérimentaux assistée par ordinateur»*, Paris : Tec&Doc Lavoisier, 1988, 368 pages.

– UdP-INRP : Actes de l'Université d'été «outils informatiques d'investigation scientifique dans l'enseignement des sciences physiques», UdP, 1993, 207 pages.

– UdP-INRP : Actes de l'Université d'été «outils informatiques d'investigation scientifique dans l'enseignement des sciences physiques», UdP, 1995, 292 pages.

Annexe 1
L'utilisation d'un tableur (Excel-Microsoft)

La commande «DROITEREG» utilise la méthode des moindres carrés pour calculer une droite qui s'ajuste au plus près des données et renvoie un tableau de valeurs.

Syntaxe : DROITEREG(*y; x; cste; stat*)

– y et x sont les séries des valeurs connues,

– *cste* est une valeur logique qui indique si la constante b doit être égale à 0 (modèle $y = ax$) ou non,

– *stat* représente une valeur logique : si *stat* est VRAI, la fonction renvoie les statistiques dans un tableau :

a	b
s_a	s_b
r^2	s_y

Note : pour utiliser une formule dite «matricielle», il faut sélectionner la zone d'affectation de 2 × 3 cellules et valider la fonction DROITEREG( ) par «Ctrl+Entrée».

Annexe 2
Schéma d'articulation maths-physique en terminale

Bernadette CAZAUX, Adeline DUCATÉ, Françoise LOPPIN, Hélène RICHOUX
Équipe du lycée Marcel Pagnol - 91200 Athis-Mons

La progression que nous avons adoptée en terminale repose sur deux points essentiels :

– introduire le questionnement sur les incertitudes *dans et par la confrontation modèle-mesures*,

– articuler le questionnement «physique» (T.P. de physique) et l'introduction de méthodes *mathématiques* (cours et T.D. de mathématiques).

Première étape : l'importance des incertitudes (T.P. de physique)

Objectif

Introduction, à l'occasion d'une activité de confrontation modèle-mesures, d'un questionnement sur la notion d'incertitude.

Sujet (exemple)

Période du pendule pesant ; l'analyse dimensionnelle conduit à une relation de proportionnalité entre T et $\sqrt{(l\,/\,g)}$ qui légitime une vérification expérimentale.

Protocole (première partie)

– Mesures de T (au chronomètre) pour différentes valeurs de l (autant que de binômes).

– Report sur un graphe (au tableau) de T en fonction de *(l/g)*.

– L'ajustement au modèle est imparfait ; il existe donc des incertitudes.

Protocole (suite)

Chaque groupe recommence sa détermination de T, trois ou quatre fois ; le graphe est complété et on fait ainsi apparaître l'étendue sur chaque détermination.

Conclusions

– À propos de la confrontation modèle-mesures : il est nécessaire de tenir compte des incertitudes.

– À propos des bornes d'étendue : on peut faire mieux avec des outils mathématiques.

Seconde étape : des outils statistiques (cours et T.D. de mathématiques)

Objectif

Faire comprendre qu'en statistiques on effectue des estimations.

Expliquer qu'une population centrée peut être caractérisée par deux grandeurs : une valeur globale, la moyenne, et une représentation de l'étendue (de la dispersion) : l'écart-type (on donne les formules).

Expliquer que lorsqu'on n'a qu'un échantillon, on ne peut avoir qu'une estimation de la moyenne et de l'écart-type et que la meilleure estimation consiste à prendre :
$m_e = \sum m_i / N$; $\sigma_e = \sqrt{\left[\sum (m_e - m_i)^2 / (N-1)\right]}$.

Application

Retour sur l'exemple du pendule ; on peut remplacer les barres d'étendue par une barre centrée sur la moyenne et de largeur égale à $2.\sigma_e$, par exemple (pour une chance donnée).

Troisième étape : cas où l'on n'a pas la possibilité de refaire chaque mesure (T.P. de physique)

Choisir l'exemple d'une modélisation par une droite à partir d'une acquisition automatique.

Activité

Faire tracer *à la main* (sur sortie imprimante) la droite qui passe au mieux *dans* l'ensemble des points. On trace donc une droite moyenne ; il existe des écarts résiduels entre chaque point et la droite.

Expliquer que ces écarts résiduels sont représentatifs des incertitudes aléatoires et que l'on peut donc avoir une information sur l'étendue en prenant le point le plus écarté ; faire noter que cela suppose bien que l'on considère tous les points comme étant équivalents du point de vue des incertitudes.

Expliquer qu'alors les valeurs a_c et b_c ainsi calculées ne sont que des valeurs approchées : ce sont des estimations des valeurs cherchées.

Utilisation d'un logiciel

Effectuer la modélisation : récrire dans une forme acceptable par le physicien l'affichage des valeurs a = xx ± uu et b = yy ± vv rendues par le logiciel.

Complément

Dire que le logiciel détermine ces paramètres par minimisation d'un «écart quadratique».

Quatrième étape : de l'écart quadratique à l'écart-type (cour et T.D. de mathématiques)

Écart quadratique

Reprendre la formule et monter l'existence d'un minimum sur le cas simple d'un modèle y = ax (J n'est alors qu'une fonction d'une seule variable).

Revenir aux aspects statistiques, et expliquer que ce choix de définition de J n'est pas indépendant de la définition de σ : montrer l'analogie ; dire qu'il existe effectivement une relation entre Jmin et σ_y.

Exercice avec un tableur : la régression linéaire permet d'obtenir a, b mais aussi σ_a, σ_b, σ_y.

Dernière étape : réinvestissement général (T.P. de physique)

Sur un autre cas, *non linéaire* (à propos de la charge d'un condensateur, par exemple) faire une modélisation-optimisation et transcrire en clair pour le physicien toutes les informations issues du calcul : Umax et RC avec leurs incertitudes, et incertitude sur chaque mesure de U.